Web Engineering @ UCL

Research Programming Social – Oct 14 2015

Ian Sillitoe (Darwin 627)

i.sillitoe@ucl.ac.uk

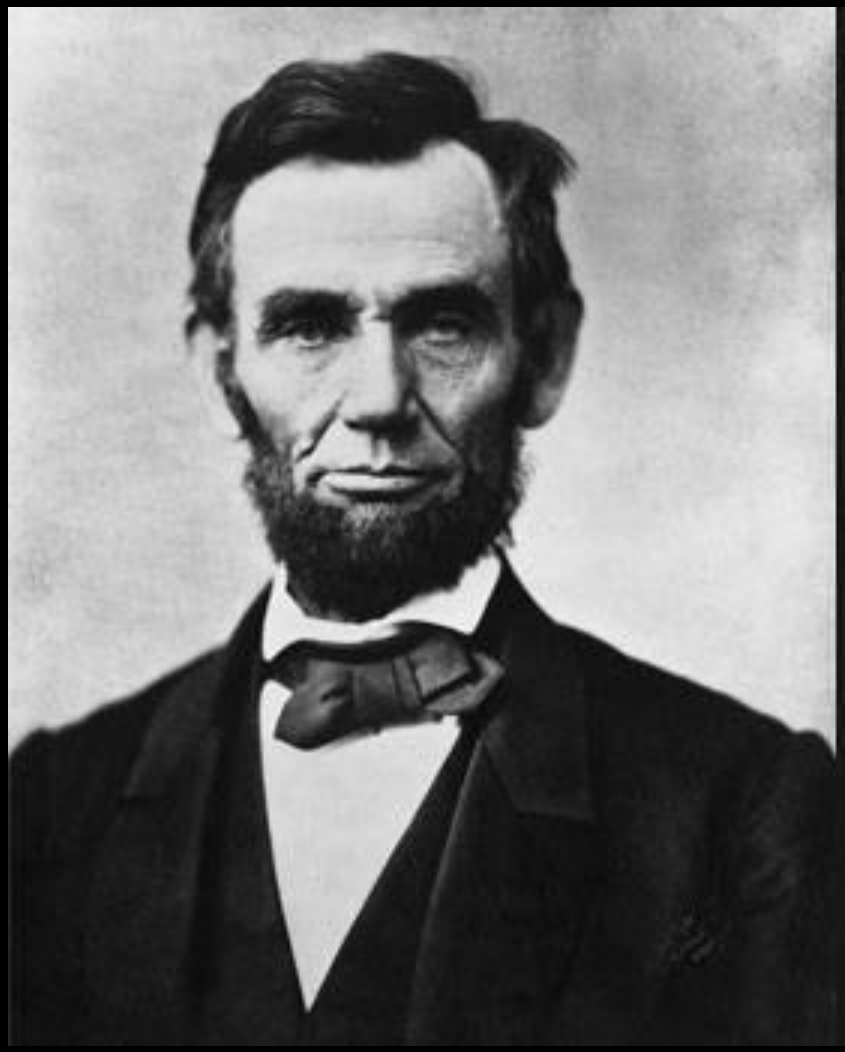
About me

- Bashing my head against various web technologies for > 20 years

- What I've learned so far:
  - There is no single, correct way to do things
  - There are lots of bad ways to do things

An expert is someone who knows some of the worst mistakes that can be made in his subject and how to avoid them.

— Werner Heisenberg —

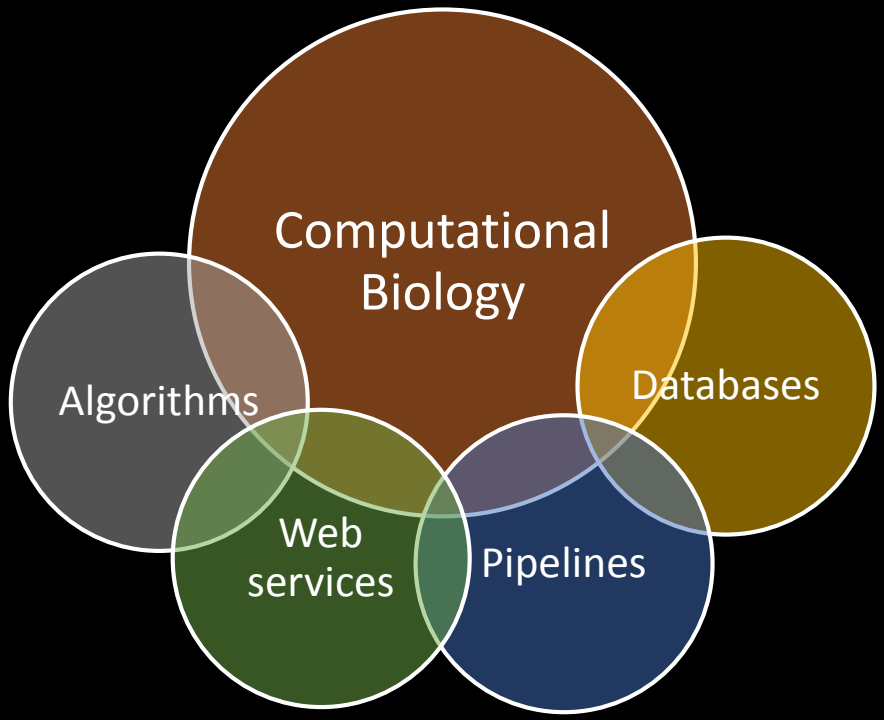


"Don't believe everything you read on the Internet just because there's a picture with a quote next to it."

—Abraham Lincoln

Orengo Group

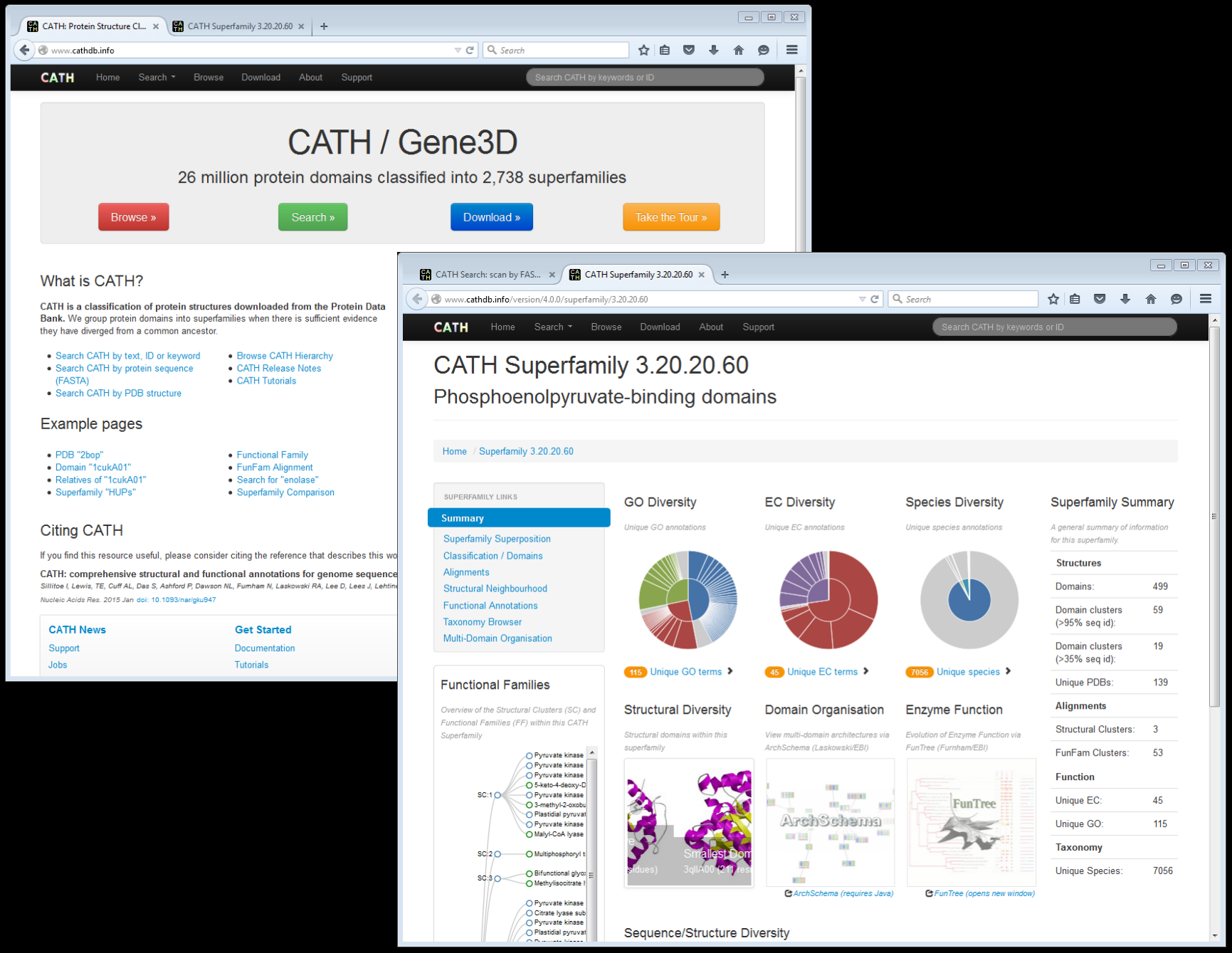
- Prof Christine Orengo
  - Room 627, Darwin
  - 6 Post docs, 6 PhD students



Linux Services

20	6	3	50
Desktops / Laptops	General Servers	Big VM Servers	Local HPC nodes

- Version Control — SVN, Git
- Web Tools — Apache, Varnish, FCGI, Solr, Tomcat
- Databases — PostgreSQL, Oracle, MySQL
- Software Dev — Perl, Python, Java, C++, Javascript
- Sys Dev — VM, Puppet, CentOS, Ubuntu


CATH: Protein Structure Cl...
CATH Superfamily 3.20.20.60
www.cathdb.info
Search
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
CATH / Gene3D
26 million protein domains classified into 2,738 superfamilies
Browse »
Search »
Download »
Take the Tour »
What is CATH?
CATH is a classification of protein structures downloaded from the Protein Data Bank. We group protein domains into superfamilies when there is sufficient evidence they have diverged from a common ancestor.
Search CATH by text, ID or keyword
Search CATH by protein sequence (FASTA)
Search CATH by PDB structure
Browse CATH Hierarchy
CATH Release Notes
CATH Tutorials
Example pages
PDB "2bop"
Domain "1cukA01"
Relatives of "1cukA01"
Superfamily "HUPs"
Functional Family
FunFam Alignment
Search for "enolase"
Superfamily Comparison
Citing CATH
If you find this resource useful, please consider citing the reference that describes this wo
CATH: comprehensive structural and functional annotations for genome sequence
Sillitoe I, Lewis, TE, Cuff AL, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees J, Lehtin
Nucleic Acids Res. 2015 Jan doi: 10.1093/nar/gku947
CATH News
Support
Jobs
Get Started
Documentation
Tutorials
CATH Search: scan by FAS...
CATH Superfamily 3.20.20.60
www.cathdb.info/version/4.0.0/superfamily/3.20.20.60
Search
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
CATH Superfamily 3.20.20.60
Phosphoenolpyruvate-binding domains
Home / Superfamily 3.20.20.60
SUPERFAMILY LINKS
Summary
Superfamily Superposition
Classification / Domains
Alignments
Structural Neighbourhood
Functional Annotations
Taxonomy Browser
Multi-Domain Organisation
GO Diversity
Unique GO annotations
115 Unique GO terms
EC Diversity
Unique EC annotations
45 Unique EC terms
Species Diversity
Unique species annotations
7056 Unique species
Superfamily Summary
A general summary of information for this superfamily.
Structures
Domains: 499
Domain clusters (>95% seq id): 59
Domain clusters (>35% seq id): 19
Unique PDBs: 139
Alignments
Structural Clusters: 3
FunFam Clusters: 53
Function
Unique EC: 45
Unique GO: 115
Taxonomy
Unique Species: 7056
Functional Families
Overview of the Structural Clusters (SC) and Functional Families (FF) within this CATH Superfamily
SC:1
SC:2
SC:3
Pyruvate kinase
Multiphosphoryl t
Bifunctional glyo
Methylisocitrate l
Citrate lyase sub
Plastidial pyruvat
Malyl-CoA lyase
Structural Diversity
Structural domains within this superfamily
Domain Organisation
View multi-domain architectures via ArchSchema (Laskowski/EBI)
ArchSchema
ArchSchema (requires Java)
Enzyme Function
Evolution of Enzyme Function via FunTree (Furnham/EBI)
FunTree
FunTree (opens new window)
Sequence/Structure Diversity

Background

- Technology
  - HTML / CSS / Javascript
- Concepts
  - MVC
  - Web application
  - Virtualisation

HTML (HyperText Markup Language)

- Provides semantic meaning (markup) to documents
  - <h1>Title</h1>
  - Paragraph
  - <table>…</table>

- HTML5 provides a rich, standardised vocabulary
  - <footer>…</footer>
  - <section>…</section>
  - <audio>…</audio>
  - <video>…</video>

"MDN HTML"

https://developer.mozilla.org/en-US/docs/Web/HTML

HTML (HyperText Markup Language)

```
<h1>Title</h1>
<p>Some text and a
 <a href="http://www.ucl.ac.uk">link</a>
</p>
```

Title

Some text and a link

HTML5

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <title>Title</title>
    <link rel="stylesheet" href="style.css">
    <script src="script.js"></script>
  </head>
  <body>
    <!-- page content -->
    <p>Some text and a
       <a href="http://www.ucl.ac.uk">link</a>
    </p>
  </body>
</html>
```

CSS (Cascading Style Sheets)

- Language to specify presentation of the document
- Separates data from presentation
- Consistency across many pages

"MDN CSS"

https://developer.mozilla.org/en-US/docs/Web/CSS

CSS (Cascading Style Sheets)

HTML

```
<h1>Title</h1>
<p>Some <span class="example">text</span> and a
 <a href="http://www.ucl.ac.uk">link</a>
</p>
```

CSS

Title

Some text and a link

CSS (Cascading Style Sheets)

HTML

```
<h1>Title</h1>
<p>Some <span class="example">text</span> and a
 <a href="http://www.ucl.ac.uk">link</a>
</p>
```

CSS

```
p .example {
        color: red;
}
```

Title

Some text and a link

Javascript

- Dynamic language
- Commonly used to negotiate interaction between user and HTML
- Standard is supported by all modern browsers
- JS Frameworks make it easier to use
  - jQuery, Underscore, Prototype, …

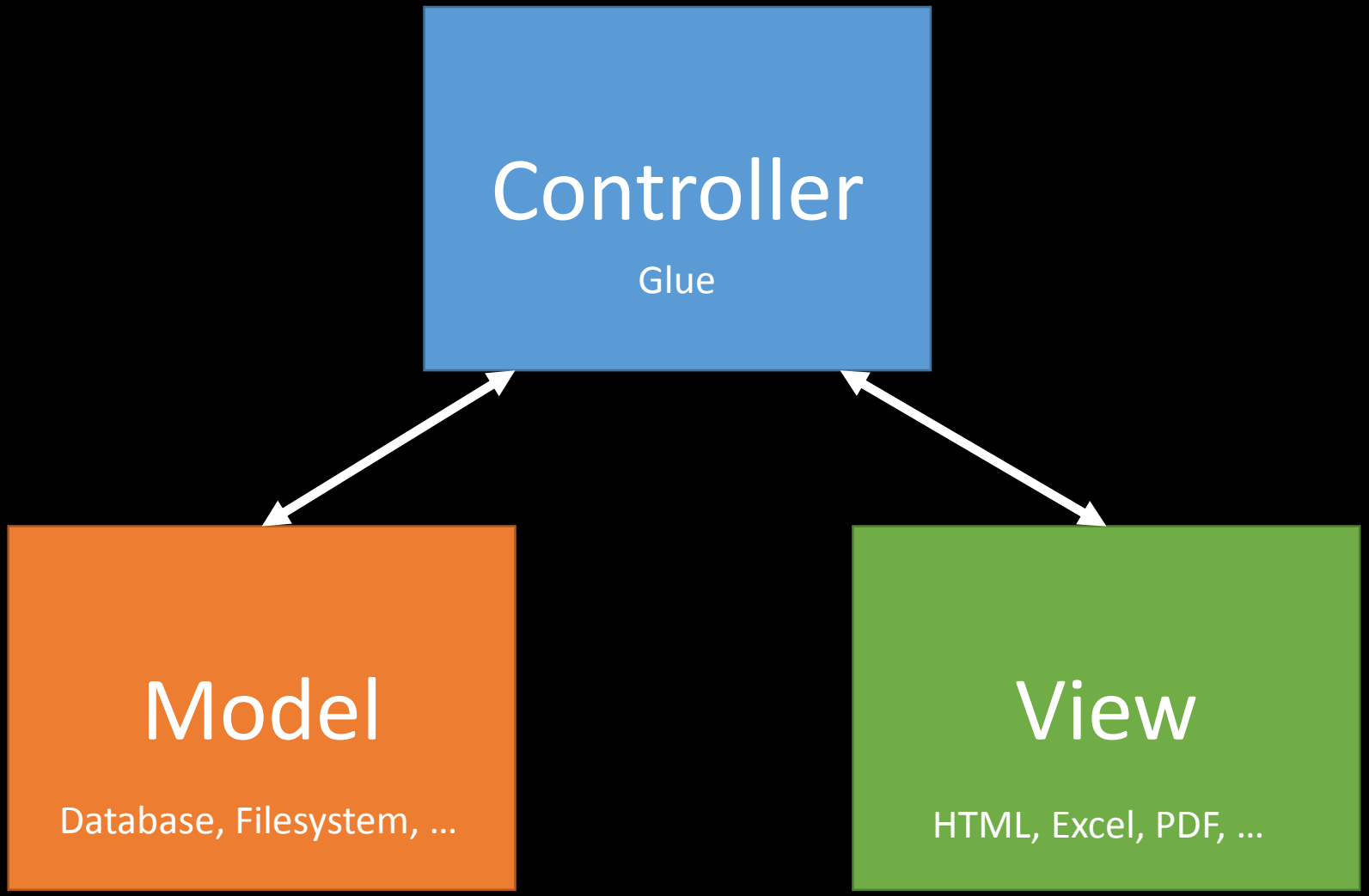
"MDN Javascript"

https://developer.mozilla.org/en-US/docs/Web/JavaScript

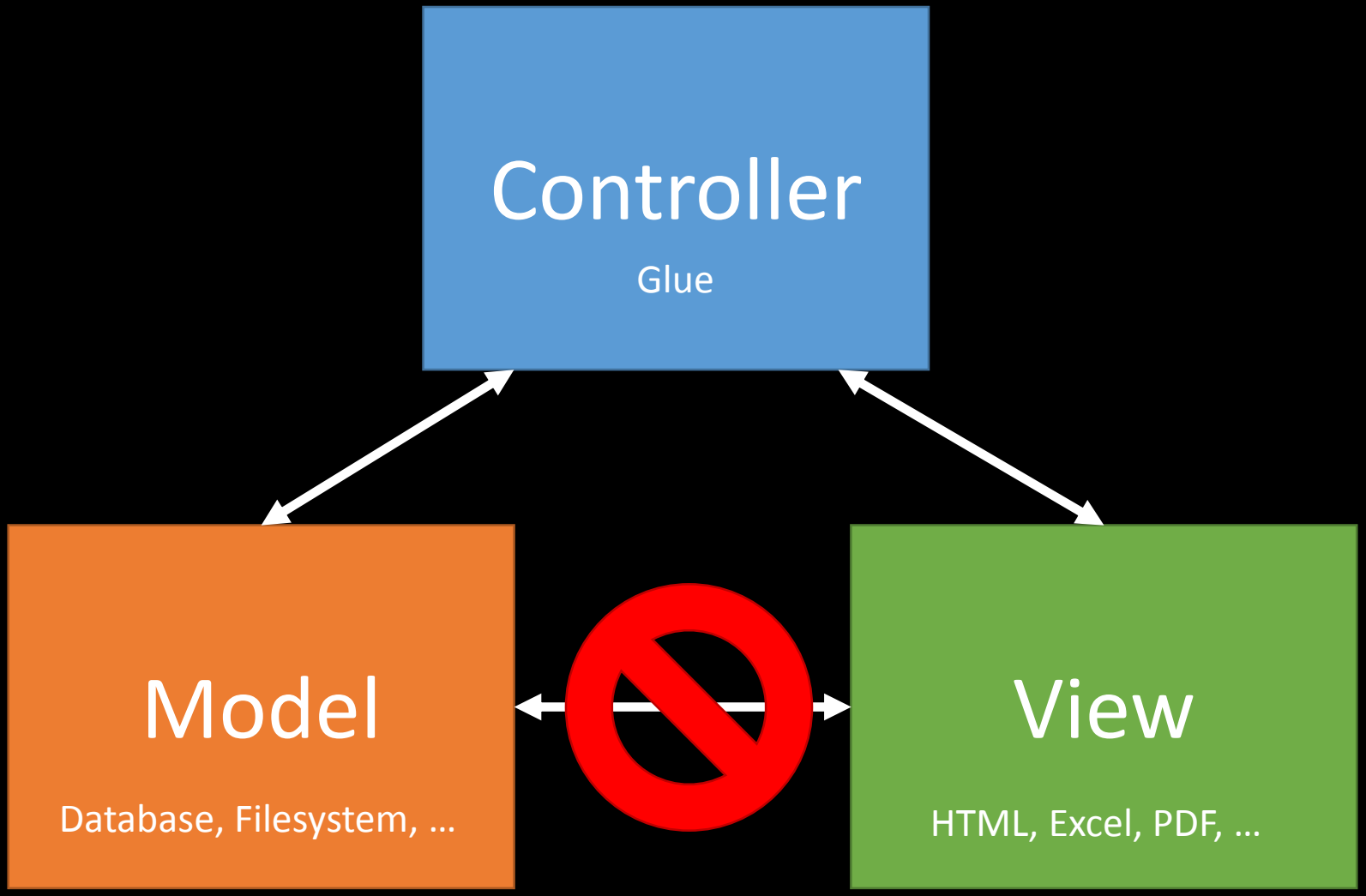
Background

- Technology
  - HTML / CSS / Javascript
- Concepts
  - MVC
  - Web application
  - Virtualisation

MVC


Controller
Glue
Model
Database, Filesystem, …
View
HTML, Excel, PDF, …

MVC


Controller
Glue
Model
Database, Filesystem, …
View
HTML, Excel, PDF, …

Crowd Computing

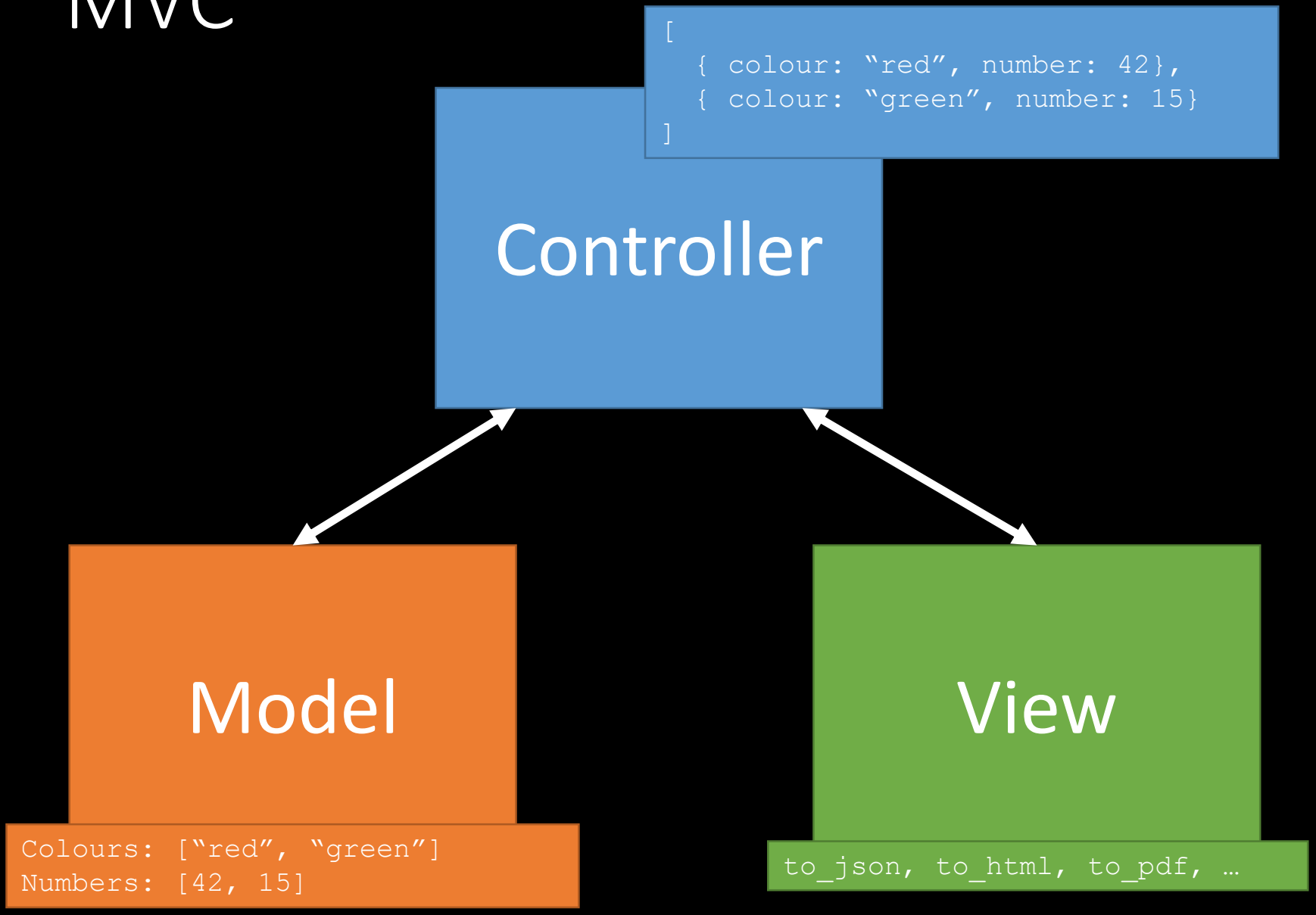
- Web Service

  `http://api.ucl.ac.uk/RandomNumberAndColour/2`

- Models
  - `RandomNumber`
  - `RandomColour`

- Views
  - `to_json`
  - `to_html`
  - `to_html( language => ‘Spanish’ )`

MVC


[
{ colour: "red", number: 42},
{ colour: "green", number: 15}
]
Controller
Model
Colours: ["red", "green"]
Numbers: [42, 15]
View
to_json, to_html, to_pdf, …

Crowd Computing

- Web Service

```
http://api.ucl.ac.uk/RandomNumberColour/2
```

- Returns

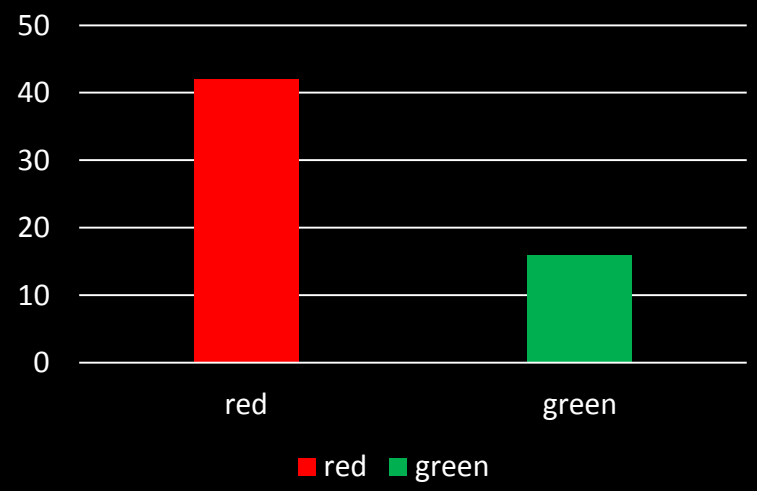
```
[
  { colour: “red”, number: 42},
  { colour: “green”, number: 15}
]
```

Crowd Computing

- Web Service

```
http://api.ucl.ac.uk/RandomNumberColour/2
```

- Returns

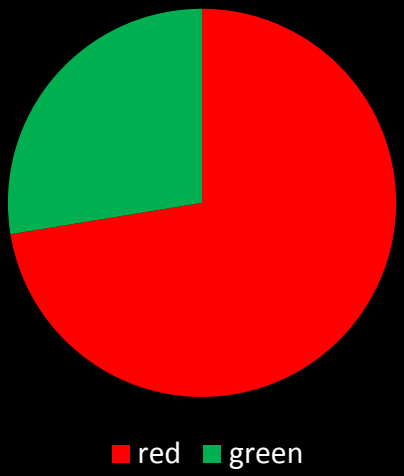


Crowd Computing

- Web Service

```
http://api.ucl.ac.uk/RandomNumberColour/2
```

- Returns



Challenge 1

- Provide web application that is:
  - Efficient
  - Robust
  - Flexible
  - Scalable
  - Maintainable

Efficient

- Use MVC web application (not CGI.pl)

- e.g.
  - Perl — **Catalyst**, Dancer, Mojolicious, …
  - Python — Django, CherryPy, …
  - Ruby — Rails, …

Robust

- Continuous Integration (CI)
  - Everything gets put into group code repository
    - code, crontabs, tests, etc
  - All code has its own tests
  - Tests run automatically all the time
  - If tests PASS - shared code gets updated
  - If tests FAIL - we get an email

- e.g. Jenkins CI, Travis CI, …

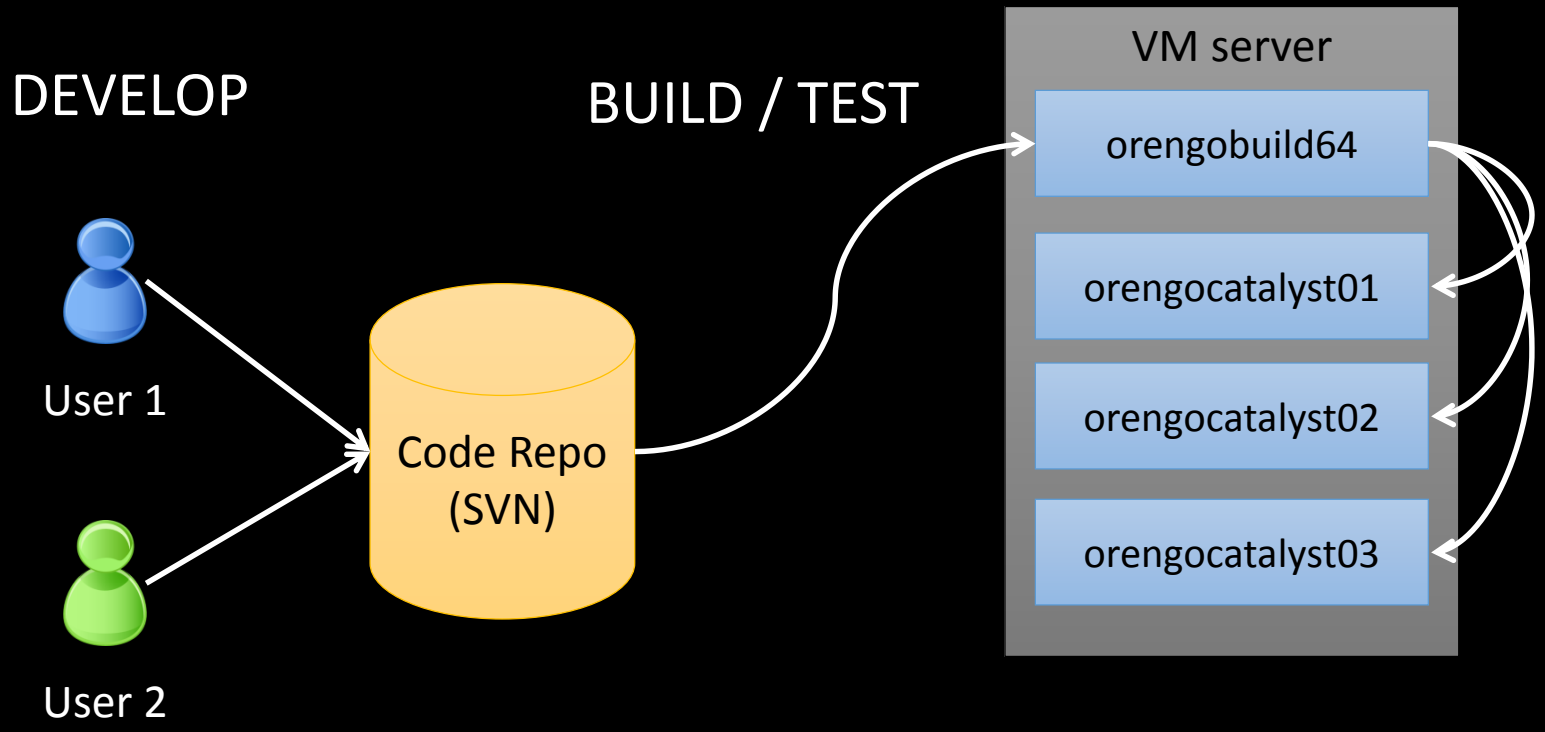
Flexible

- Lots of (good) tests gives you…
  - Confidence to make changes
  - Confidence to accept changes from others
- Continuous Integration lets you…
  - See your own changes in “beta” sites
- MVC web applications make it easy to…
  - Create powerful, interactive web pages

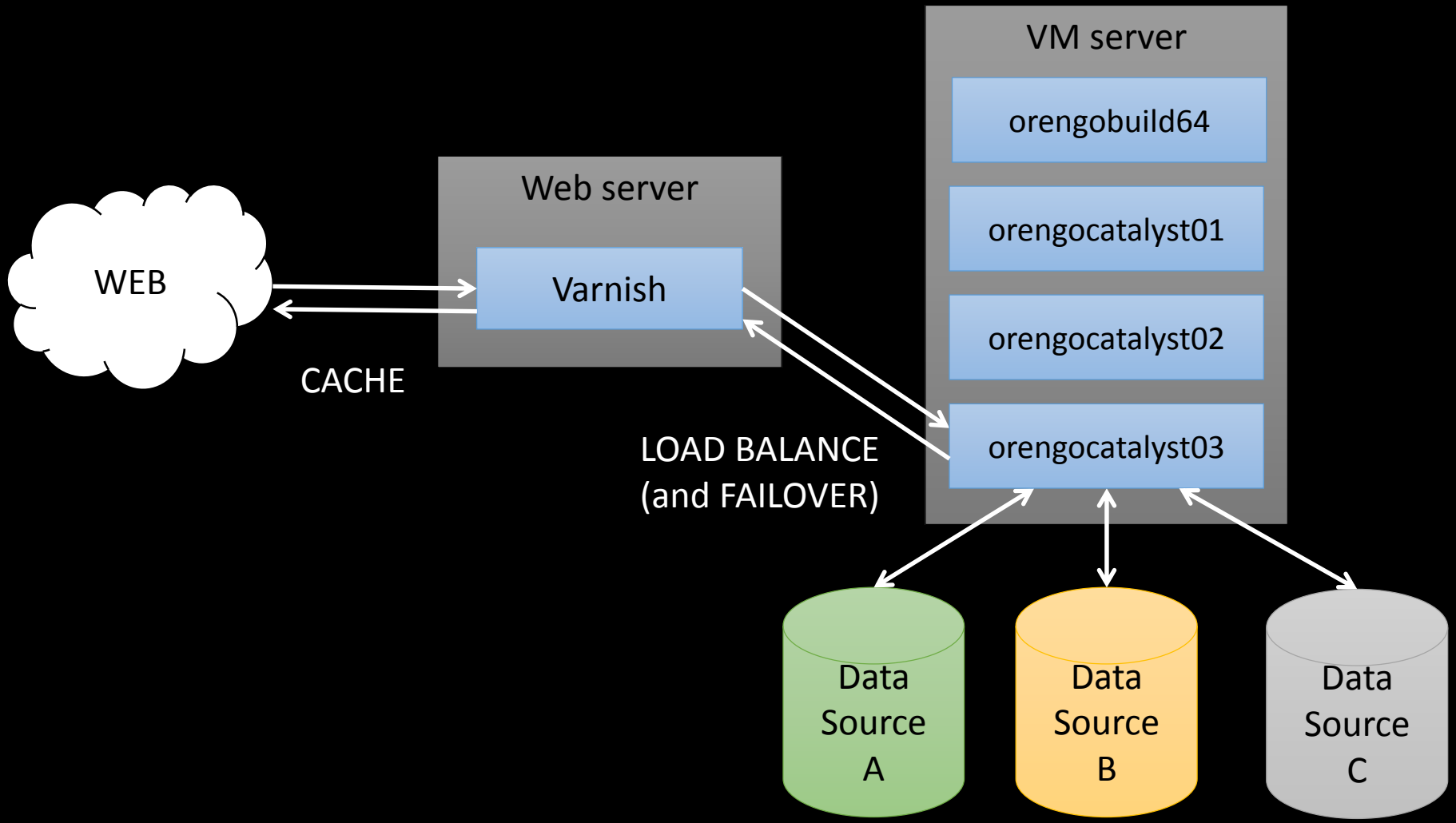
Scalable

- Move web code to Virtual Machine (VM)
  - SVN all common code in central repo
  - Puppet maintains machine requirements
  - Varnish web cache, load balance, failover

Develop, Build / Test, Deploy



Cache, Load Balance, MVC



Challenge 2

- Monitor SGE jobs (qstat without the typing)

- https://github.com/sillitoe/sge-monitor